



# *Providing access to your data*

---

Ruth Duerr

National Snow and Ice Data Center, University of Colorado





# Outline

---

- Why provide access?
- Rights
- Handling Sensitive Data
- Determining your audience
- Mechanisms
  - for providing data access
  - for advertising the existence of your data



# Arguments for Making Data Accessible

---

- ***Agency requirements.*** Many agencies require it.
- ***Legal considerations.*** Data produced through public investments are public goods and should properly be in the public domain.
- ***Socio-economic considerations.*** Open online access is the most efficient way to disseminate data in order to maximize their value to the public.
- ***Ethical considerations.*** The public has already paid for the production of data, and thus should be able to access the data.
- ***Good governance considerations.*** Transparency of data and scientific activities are undermined by restricting access to and use of public data.



# Funding Agency Requirements

---

- National Science Foundation policy on “Dissemination and Sharing of Research Results”:
  - “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.” [http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag\\_6.jsp#VID4](http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4)
- The NSF GeoScience directorate specifies data management requirements for Ocean Science, Earth Science, and Atmospheric and Geospace Sciences.
  - Data management planning requirements differ amongst the sciences
  - A number of data archives are suggested for each science
  - See <http://www.nsf.gov/geo/geo-data-policies/index.jsp> for details



# Funding Agency Requirements

---

- NASA “Data & Information Policy”:
  - “NASA commits to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available.
  - There will be no period of exclusive access to NASA Earth science data...
  - NASA will make available all NASA-generated standard products along with the source code for algorithm software, coefficients, and ancillary data used to generate these products.
  - All NASA Earth science missions, projects, and grants and cooperative agreements shall include data management plans to facilitate the implementation of these data principles.” <http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/>



# Funding Agency Requirements

---

- NOAA
  - “NOAA recognizes that open and unrestricted dissemination of high quality publicly funded information, as appropriate and within resource constraints, is good policy and is the law.” <http://www.noaa.gov/partnershippolicy/>
  - Proposed Data Policy: “All environmental data and information collected or created under NOAA grants or cooperative agreements awarded after [date] must be made visible, accessible, and independently understandable to users in a timely manner, except where limited by law, regulation, policy or by security requirements.” (Guch, 2011)
- USGS “Information Policy and Instructions”:
  - “USGS-authored or produced data and information are considered to be in the U.S. public domain.” [http://www.usgs.gov/laws/info\\_policies.html](http://www.usgs.gov/laws/info_policies.html)



# Reality

---

- Sharing data
  - It can be difficult to share data with people who aren't familiar with your project
  - Some data might be sensitive, and thus not shareable (or only partially shareable)
  - It might be difficult to determine who is responsible for sharing data, and who has permission to share data within a project



# Copyright and Data

---

- Not copyrightable
  - Facts:  $2+2 = 4$
  - Measurements: Temperature = 20 C
- Copyrightable
  - Original selection and arrangement of data
  - Metadata: documentation and descriptions of data or processing steps
- Copyrights are assigned by default, i.e. without having to officially file for them.
  - Ex. lab notebooks or dissertations are copyrighted without researchers filing for them.
- Through employment contracts/agreements, copyrights for research products are often owned by employers (e.g. universities), not creators.





# Making Data Public

---

- Copyright licenses for making data public exist:
  - Free/Open Licenses such as GNU General Free Documentation License (GFDL)
  - Creative Commons licenses

BUT

- Such copyright licenses will inevitably complicate data sharing and integration efforts of any large scale
  - There is an amorphous line between what can and cannot be copyrighted within databases
  - Different licenses have different usage requirements, e.g. querying across 10 databases may return results with 10 different usage licenses



# Appropriate Data Use

---

- Putting data in the public domain is the solution
  - Eliminates data use restrictions
  - Enables data integration
  - Encourages non-legal means for resolving problems
- Communities should develop norms as to how data should be made available, used, and attributed
  - Check with data centers for data use and attribution policies
  - Work with collaborators to ensure that the usage and attribution of others' data meets with community accepted practices
  - Examples:
    - International Polar Year
    - Seismology data
    - Federation of Earth Science Information Partners (ESIP)



# Reasons why data may be sensitive?

---

- Need to protect the confidentiality of human subjects
- Need to protect the rights of knowledge holders (e.g., local or traditional knowledge)
- Need to prevent harm (e.g., endangered species, sacred sites)



# Methods of providing access to sensitive data

---

- Limit access to authorized individuals or roles
  - Enable access within a facility only accessible by authorized persons.
  - Limit access to authorized individuals with password-protection.
  - Establish a data enclave without capabilities to copy or transmit data.
- Provide options for authorizing access to sensitive data
  - Require evidence of approval or expertise.
  - Require users to sign a confidentiality, non-disclosure, or data use agreement.
  - Require approval of data protection plan from potential users.
  - Request an exception or waiver of restrictions under certain conditions.
  - Request clearance for specific individuals to access sensitive data.
  - Request declassification of sensitive data from relevant authority.
- Modify the data to protect sensitive portions
  - Change aspects of data that are sensitive (recode, generalize)
  - Remove or obscure the portions of data that are sensitive (redact)



# Ways of making your data accessible

---

- Do it yourself!
- Find an archive or repository to hold your data (i.e., transfer the responsibility to them!)



# Issues with rolling your own

---

- Do you really have time to...
  - be a good data manager?
  - provide good user support?
- Do you really want to...
  - spend your grant money on the hardware, software, and personnel needed?
- And what do you do when your grant runs out?



# Why determine your audience?

---

- Because the audience determines everything!
  - What a good data format would be
  - The amount of detail and language to use in describing the data
  - What kind of tools you should be prepared to support
  - etc.

# An example: Data of public interest

```
rduerr@sidads:/ftp/pub/DATASETS/parca/nsidc0218_melt/melt_raw/1981> more 1981209smr.meltpts
```

```
3 21
4 21
.
```

```
rduerr@sidads:/ftp/pub/DATASETS/parca/nsidc0218_melt/melt_raw/1981> more 1981209smr.meltpts
```

```
3 21
4 21
4 22
4 28
4 29
```

```
8 21
9 19
9 20
9 31
10 18
10 19
10 20
10 31
11 17
11 18
11 19
11 33
12 17
12 18
12 19
12 34
12 35
12 37
13 17
13 18
13 36
13 37
13 38
13 39
```

```
--More--(7%)
```

What we make available:

- provided ascii data files
- gridded binary data files
- annual climatology files in binary and geotiff
- climatology for the entire series in binary and geotiff







# Finding an archive

---

- Your funding source may have mandated or recommended a repository
- There may be an appropriate discipline specific repository
  - <http://datacite.org/repolist>
  - [http://oad.simmons.edu/oadwiki/Data\\_repositories](http://oad.simmons.edu/oadwiki/Data_repositories)
  - <http://d2c2.lib.purdue.edu/OtherRepositories.php>
- Your institution may have a repository perhaps through a library or computing center



# Why advertise your data?

---

- How else are potential users going to find it?
- If no-one knows that it exists can you really say that you are sharing it?

# Advertising Scenarios

- You've deposited your data in an archive
  - Advertising your data is their problem (mostly)!
- Word of mouth advertising
- Advertising your data in a directory, registry, or clearinghouse
- Advertising your data on the web

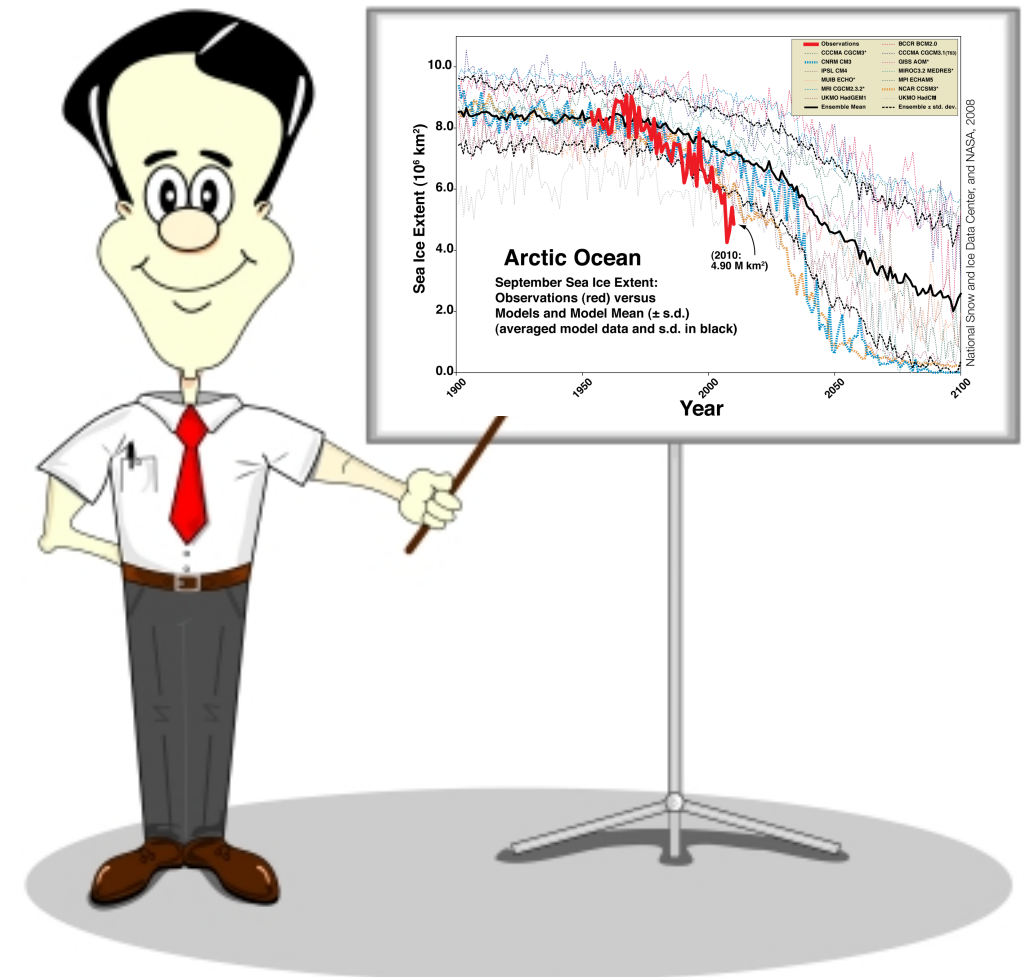


Image: Grant Cochrane / FreeDigitalPhotos.net



# Discovery Level Metadata

---

- The information needed to determine whether a data set meets a users needs.

***What:*** Title of Data Set and Keywords Describing the Data Set

***Why:*** Description and Purpose of the Data Set

***When:*** Temporal Coverage of the Data Set

***Who:*** Data Set Creator and Contact

***Where:*** Geographic Extent and Location of Data Set Coverage

***How:*** How the Data Set was Created and How to Access the Data



# What is a Metadata Clearinghouse?

---

- A metadata clearinghouse is a repository of metadata records and sometimes data
- Clearinghouses make metadata records easy to find
- Examples:
  - Global Change Master Directory (<http://gcmd.nasa.gov/>)
  - Data.gov (<http://www.data.gov/>)
  - Climate Change Database Clearinghouse ([http://ccrm.vims.edu/climate\\_change/index.html](http://ccrm.vims.edu/climate_change/index.html))



# Global Change Master Directory (GCMD)

---

- The GCMD enables the scientific community to discover and access over 25,000 Earth science data sets and services.
- The GCMD uses the Directory Interchange Format (DIF), a discovery level metadata standard that is used by NASA and the Committee on Earth Observation Satellites (CEOS).
- The DIF contains several fields, allowing users to easily find and determine the data set and/or service that is useful to their needs.



# Publishing Metadata

---

- docBUILDER is a free, online tool for creating, updating, and publishing metadata entries to the GCMD.
  - Provides a “visual” check-list of required, highly recommended, and recommended fields to populate.
  - Performs validation (checks controlled fields and syntax).
  - Allows for templates when creating entries with repeatable information.
- Users can register for an account at <https://users.eosdis.nasa.gov>
- Users can access docBUILDER at <http://gcmd.nasa.gov/User/authoring.html>
- DIF entries provide a direct link (i.e. [http://gcmd.nasa.gov/getdif.htm?GES\\_DISC\\_AIRI2CCF\\_NRT\\_V005](http://gcmd.nasa.gov/getdif.htm?GES_DISC_AIRI2CCF_NRT_V005)) to metadata that can be added to your website.



# Using docBUILDER

Validate Your Document  
and/or Get Help

Submit Your  
Document  
to the GCMD

Field is  
checked  
when  
completed

The screenshot shows the docBUILDER interface with a menu bar (FILE, DOCUMENT, HELP) and a 'Submit to GCMD' button. A list of metadata fields is displayed, categorized by requirement level: Required (orange square), Highly Recommended (yellow square), and Recommended (green square). The 'Entry ID' field is checked. A note at the bottom states: 'Note: This document is automatically saved and can be retrieved up to 90 days using the Entry ID (Document Identifier) "ESIP". Please complete the document and submit it for publication within that period.' A legend at the bottom explains the requirement levels.

Requirement Level	Field Name
Required	Entry ID
Required	Entry Title
Required	Science Keywords
Required	ISO Topic Category
Required	Data Center
Required	Summary
Highly Recommended	Data Set Citation
Highly Recommended	Personnel
Highly Recommended	Related URL
Highly Recommended	Instrument
Highly Recommended	Platform
Highly Recommended	Temporal Coverage
Highly Recommended	Paleo-Temporal Coverage
Highly Recommended	Spatial Coverage
Highly Recommended	Location
Highly Recommended	Data Resolution
Highly Recommended	Project
Highly Recommended	Quality
Highly Recommended	Access Constraints
Highly Recommended	Use Constraints
Highly Recommended	Distribution Information
Highly Recommended	Data Set Language
Highly Recommended	Data Set Progress
Recommended	Ancillary Keyword
Recommended	Originating Center
Recommended	Multimedia Sample
Recommended	Publication/Reference
Recommended	Parent DIF
Recommended	IDN Node
Recommended	DIF Creation Date
Recommended	Last DIF Revision Date
Recommended	DIF Revision History
Recommended	Future DIF Review Date
Recommended	Privacy Status

Note: This document is automatically saved and can be retrieved up to 90 days using the Entry ID (Document Identifier) "ESIP". Please complete the document and submit it for publication within that period.

Legend: ■ = Required ■ = Highly Recommended ■ = Recommended

Populate Metadata Fields





# Why isn't posting a web page enough?

Google

rainfall data over north dakota

×

🔍

Everything

Images

Videos

News

More >

⚙️

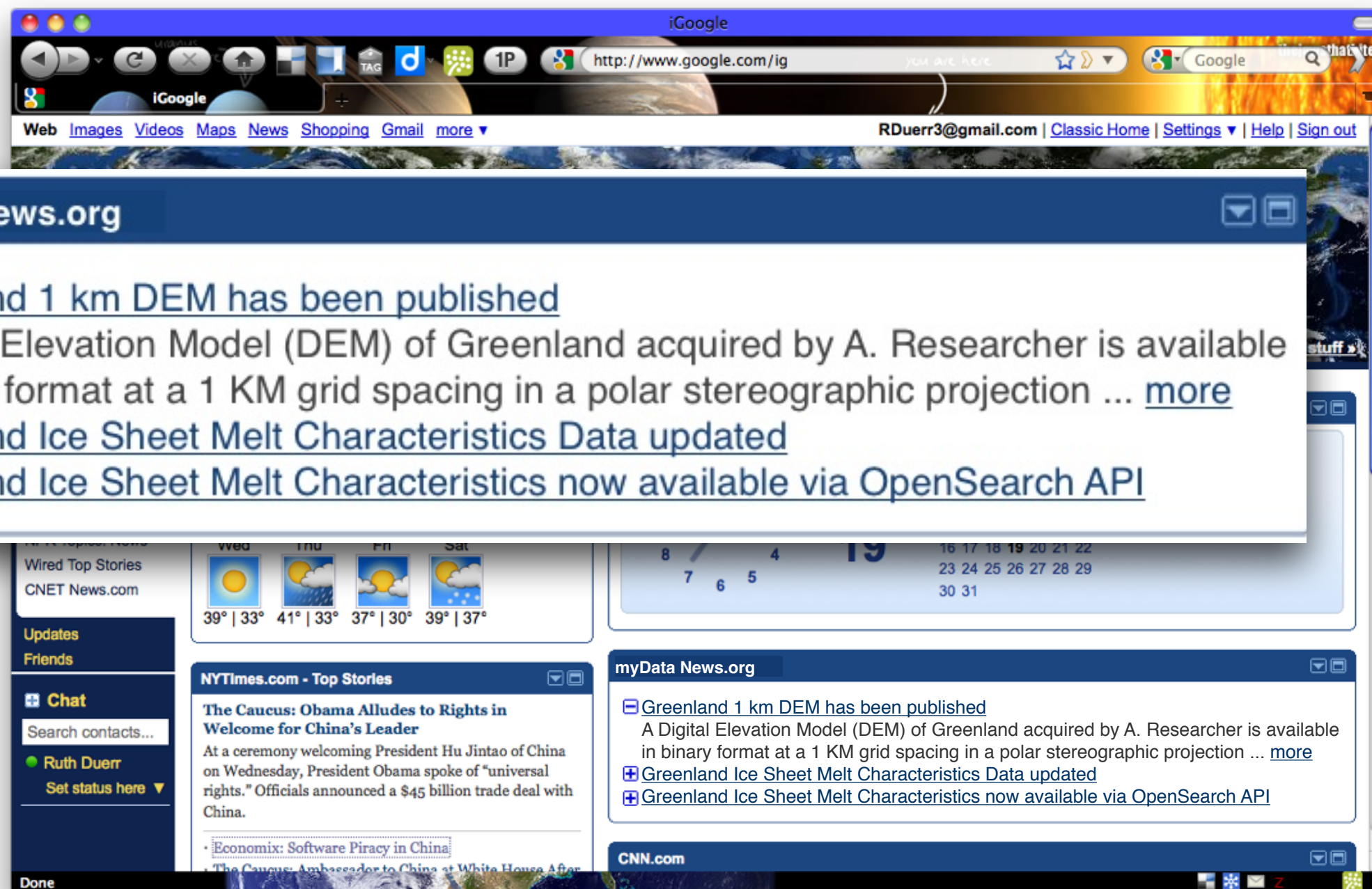
**Rain Fall Totals By City | Ask.com**  
[www.ask.com/Rain+Fall+Totals+By+City](http://www.ask.com/Rain+Fall+Totals+By+City)  
View **Rain Fall** Totals By City; Get Answers Now on Ask.com! Ad ⓘ

**Red River basin precipitation - Climatology - Division of Waters ...**  
[www.dnr.state.mn.us](http://www.dnr.state.mn.us) > ... > [Ground Water](#) > [Climatology](#)  
All or part of five climate divisions in **North Dakota**, three in Minnesota, and one in ...  
**Precipitation data** from these five climate divisions are used to provide the ... However, **over** these sixty years, a great deal of variability can be observed.

**North Dakota Monthly Precipitation**  
[www.ndsu.edu/ndsco/precip/monthly/2011.html](http://www.ndsu.edu/ndsco/precip/monthly/2011.html)  
north dakot state climate office logo. AASC Recognized Climate Office ... 2011 **North Dakota** Monthly **Precipitation**. Click Snow Tab for 2011-2012 Monthly Snow ...

**Coop Observer Network - ND State Water Commission**  
[www.swc.nd.gov/4dlink9/4dcgi/.../Coop%20Observer%20Network](http://www.swc.nd.gov/4dlink9/4dcgi/.../Coop%20Observer%20Network)  
nd.gov - The Official Portal for **North Dakota** State Government ... Precipitation Download -

# Publishing metadata to the web

A screenshot of a web browser window showing a Google search for 'myData News.org'. The browser is iGoogle, with the address bar showing 'http://www.google.com/ig'. The search results for 'myData News.org' are displayed, showing a list of news items. The first item is 'Greenland 1 km DEM has been published', followed by 'Greenland Ice Sheet Melt Characteristics Data updated' and 'Greenland Ice Sheet Melt Characteristics now available via OpenSearch API'. The browser also shows a sidebar with 'Wired Top Stories' and 'CNET News.com', and a 'myData News.org' widget at the bottom right.

myData News.org

- ☐ [Greenland 1 km DEM has been published](#)  
A Digital Elevation Model (DEM) of Greenland acquired by A. Researcher is available in binary format at a 1 KM grid spacing in a polar stereographic projection ... [more](#)
- ☐ [Greenland Ice Sheet Melt Characteristics Data updated](#)
- ☐ [Greenland Ice Sheet Melt Characteristics now available via OpenSearch API](#)

Wired Top Stories  
CNET News.com

Updates  
Friends

Chat  
Search contacts...

Ruth Duerr  
Set status here ▼

NYTimes.com - Top Stories

**The Caucus: Obama Alludes to Rights in Welcome for China's Leader**

At a ceremony welcoming President Hu Jintao of China on Wednesday, President Obama spoke of "universal rights." Officials announced a \$45 billion trade deal with China.

Economix: Software Piracy in China

The Caucus: Ambassador to China at White House After

myData News.org

- ☐ [Greenland 1 km DEM has been published](#)  
A Digital Elevation Model (DEM) of Greenland acquired by A. Researcher is available in binary format at a 1 KM grid spacing in a polar stereographic projection ... [more](#)
- ☐ [Greenland Ice Sheet Melt Characteristics Data updated](#)
- ☐ [Greenland Ice Sheet Melt Characteristics now available via OpenSearch API](#)

CNN.com



# How to cast your data collection

- Fill out a web form
- Save it to your website
- Add its link to your site

A screenshot of a web browser showing the 'Data Set' creation form on the National Snow and Ice Data Center (NSIDC) website. The browser's address bar shows the URL 'http://testsnowtest.org/libre/apps/cast/dataset/'. The page header includes the NSIDC logo, a search bar, and navigation links like 'Data', 'Programs & Projects', 'Science', 'Publications', 'News & Events', and 'About'. The main content area is titled 'Libre Data Casting, Services, & Applications' and features a sidebar with 'Home', 'Share', 'Discover', and 'About'. The form itself has tabs for 'Data set', 'Citation', 'Attributes', 'Coverage', 'Contacts', 'Rights', and 'Data Cast'. The 'Data set' tab is active, showing fields for 'Entry ID', 'Data set title', 'Data set progress' (set to 'Planned'), 'Data set summary', and 'Data set language' (set to 'English'). A 'Next' button is at the bottom of the form. The footer includes the Colorado University of Colorado at Boulder logo, contact information, and logos for NASA, NSF, and the Polar Information Commons. A 'WICKET AJAX DEBUG' button is visible in the bottom right corner.

<http://nsidc.org/libre/share/collectioncaster.html>

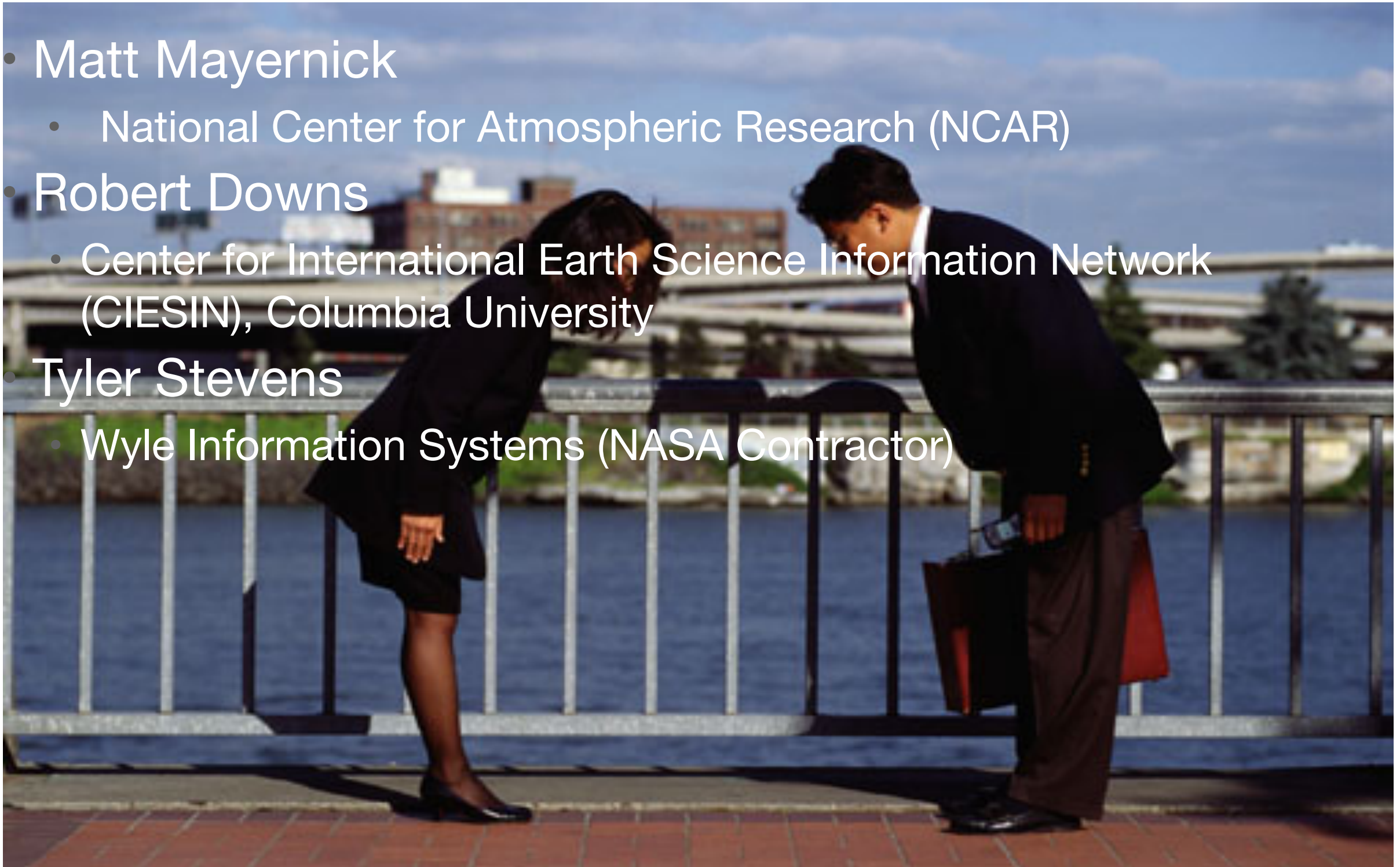




# Acknowledgements

---

- Matt Mayernick
  - National Center for Atmospheric Research (NCAR)
- Robert Downs
  - Center for International Earth Science Information Network (CIESIN), Columbia University
- Tyler Stevens
  - Wyle Information Systems (NASA Contractor)





# References and Resources

---

- Where to go for more information
- **Best Practices for Sharing Sensitive Environmental Geospatial Data**
  - [http://www.geoconnections.org/publications/Key\\_documents/Sensitive\\_Env\\_Geo\\_Data\\_Guide\\_EN\\_v1.pdf](http://www.geoconnections.org/publications/Key_documents/Sensitive_Env_Geo_Data_Guide_EN_v1.pdf)
- **Computer Security Act of 1987**
  - [http://csrc.nist.gov/groups/SMA/ispab/documents/csa\\_87.txt](http://csrc.nist.gov/groups/SMA/ispab/documents/csa_87.txt)
- **Confidentiality Issues in Geospatial Data Applications**
  - <http://sedac.ciesin.columbia.edu/confidentiality/>
  - **Dealing With Sensitive Data**
  - <http://www.dcc.ac.uk/events/research-data-management-forum/rdmf4-dealing-sensitive-data>
- **ICPSR – Restricted Data**
  - <http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/restricted/>
- **Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns**
  - <http://www.fgdc.gov/policyandplanning/Access%20Guidelines.pdf>



# References and Resources

---

- Apache Software Foundation: Licenses <http://www.apache.org/licenses/>
- BSD License <http://www.freebsd.org/copyright/license.html>
- Creative Commons <http://creativecommons.org/>
- Gnu Public Licenses (GPL & LGPL) <http://www.gnu.org>
- Open Data Foundation <http://www.opendatafoundation.org/>
- United States Copyright Office <http://www.copyright.gov>



# References and Resources

---

- Federal Geographic Data Committee (FGDC). Geospatial Metadata. <http://www.fgdc.gov/metadata>.
- National Aeronautics and Space Administration (NASA). 2010. “Directory Interchange Format (DIF) Writer's Guide”. Global Change Master Directory. <http://gcmd.nasa.gov/User/difguide/>
- National Aeronautics and Space Administration (NASA). Global Change Master Directory (GCMD). <http://gcmd.nasa.gov/>
- National Aeronautics and Space Administration (NASA). docBUILDER: <http://gcmd.nasa.gov/User/authoring.html>
- Taylor, Mark. 2000. “Developing Spatial Data Infrastructures – The SDI Cookbook”. Global Spatial Data Infrastructure Association. <http://www.gsdi.org/pubs/cookbook/chapter03a.html>.